Electrostatic Binding Energy Calculation Using the Finite Difference Solution to the Linearized Poisson–Boltzmann Equation: Assessment of Its Accuracy

JIAN SHEN*

Marion Merrell Dow Research Institute, 2110 East Galbraith Road, P.O. Box 156300, Cincinnati, Ohio 45215-6300

JOHN WENDOLOSKI

Amgen Inc., Mail Stop 14-2-A-210, Thousand Oaks, California 91340

Received 1 September 1994; accepted 12 June 1995

ABSTRACT

A full account of how to calculate the electrostatic binding energy using the finite difference solution to the linearized Poisson-Boltzmann equation (FDPB) for protein–ligand systems is described. The following tests show that the statistical and systematic errors due to discrete grid representation of molecular shape and charges amount to about 1% and 5% of calculated binding energy difference, respectively. The greater accuracy results from a three-stage error cancellation: first in ΔG_s, then $\Delta\Delta G_s$, and finally $\Delta\Delta G_{ele}$. We conclude in this study that the intrinsic error of FDPB is mostly canceled in computing binding energy differences. Among the parameters examined, the partial charge, dielectric constant, and radius of solvent can influence the calculated results most.

The quantitative calculation of binding free energy between a ligand and a protein is an important step toward understanding molecular recognition. With more and more 3D structures of protein–ligand complexes experimentally determined, accurate and fast binding energy calculation is indispensable to validate mechanistic models and to direct the design of artificial molecules for better binding. The statistical perturbation methods, such as free energy perturbation and thermodynamic integration, are considered to be capable of estimating the binding energy.[1,2] However, frequently reported shortcomings and difficulties in applications limit their usage in

* Author to whom all correspondence should be addressed.

CCC 0192-8651 / 96 / 030350-08

practice.[3,4] In addition, these methods are currently unable to cope with the speed of synthetic efforts in pharmaceutical discoveries.

In a short report,[5] an alternative approach using the finite difference solution to Poisson-Boltzmann equation (FDPB) was presented for two protein–ligand systems. It has been demonstrated that the FDPB method is a powerful tool in calculating electrostatic binding free energy differences and in elucidating noncovalent binding mechanisms. By incorporating empirical solvation entropy into the binding energy calculation, we recently extended the application for the third molecular system to include some hydrophobic modifications.[6]

Due to the grid representation of a molecular system, the FDPB method is considered to have intrinsic error in electrostatic calculations.[7–10] Gilson et al.[11] found that an error of 15% for electrostatic potential calculated by FDPB could exist at the binding distance (3 Å) around an ion. Using dielectric boundary smoothing, Davis and McCammon[12] reported a reduction of error in calculated potential of less than 10%. They also showed a much smaller error for Born solvation energy at fine grid spacing. Bruccoleri[13] demonstrated that the uniform charge model can greatly reduce the self-energy, which is a source of intrinsic error in FDPB calculations. In terms of electrostatic potentials outside of a molecule and solvation energies, however, this new model is similar to the widely used point charge model. Recently, Shen and Quiocho[5] gave a conservative error estimate of 0.2 kcal/mol for the calculated binding energy differences based on a preliminary test. Because the accuracy of the calculation and the effects of various parameters on the calculated binding energy differences have not been thoroughly studied, it seems necessary to carry out a systematic error assessment to clarify the accuracy of binding energy calculated using the FDPB method.

In this article we first give a full account of how to calculate the electrostatic binding energy for protein–ligand systems because the approach has a direct impact on the error reduction. We then investigate the effects of varying grid spacing and grid origin on the calculated binding energies. The sensitivity of parameters for solvent and solute (such as partial charges and radius of the solvent probe) to the binding energy are also examined. We conclude in this study that the intrinsic error of FDPB is mostly canceled in computing binding energy differences.

Method

BINDING ENERGY DIFFERENCE

For a noncovalent binding process, $L + P = LP$, the binding free energy, ΔG, can be partitioned into electrostatic and nonelectrostatic components, ΔG_{ele} and ΔG_n, respectively:

$$\Delta G = \Delta G_{ele} + \Delta G_n$$

Thus, the binding energy difference between a reference system and a modified system is

$$\Delta\Delta G = \Delta\Delta G_{ele} + \Delta\Delta G_n$$

We have developed a method to estimate ΔG_n based on empirical solvation entropy.[6] In this work we focus solely on ΔG_{ele} that is further decomposed into solvation (ΔG_s) and assembly (ΔG_a) energies, shown in the scheme

$$\begin{array}{ccc} L + P\ (\varepsilon = 78) & \xrightarrow{\Delta G_{ele}} & LP\ (\varepsilon = 78) \\ \Big\downarrow -\Delta G_{s,f} & & \Big\uparrow \Delta G_{s,b} \\ L + P\ (\varepsilon = 3) & \xrightarrow[\Delta G_a]{} & LP\ (\varepsilon = 3) \end{array}$$

$$\Delta G_{ele} = \Delta G_{s,b} - \Delta G_{s,f} + \Delta G_a$$

$$\Delta\Delta G_{ele} = \Delta\Delta G_{s,b} - \Delta\Delta G_{s,f} + \Delta\Delta G_a$$

where ε is the dielectric of environment in the calculation. In this approach, two solvation energies are obtained from four electrostatic calculations using FDPB and the assembly energy is analytically calculated using Coulomb's law. Due to the solute dielectric (i.e., three in our work), an identical dielectric ($\varepsilon = 3$) is used for gas phase to simplify the calculation of ΔG_a.

In principle, one need only calculate two electrostatic energies for LP ($\varepsilon = 78$) and $L + P$ ($\varepsilon = 78$) to obtain ΔG_{ele} and then $\Delta\Delta G_{ele}$. Although one can save 50% of the central processing unit (CPU) time (which is not a critical factor in FDPB applications) in this approach, most of the important information is lost except the numerical value of $\Delta\Delta G_{ele}$. In contrast, the solvation energies and assembly energies generated in our approach have their own physical meaning and valid numerical values. The analysis of these energies leads to mechanistic insight into the macromolecular binding interactions and there could suggest new directions to improve the binding.[6]

ELECTROSTATIC ENERGY CALCULATION

The electrostatic energy (E) of a molecular system is calculated through

$$E = 0.5\Sigma q_i\phi$$

where q_i is a charge and the potential, ϕ, is solved using the linearized Poisson-Boltzmann (PB) equation:

$$-\nabla \cdot \varepsilon\nabla\phi + \varepsilon k^2\phi = \rho$$

First, the coordinates of a complex are divided into three sets, **1**, **2** for a protein and **3** for a ligand, shown in Figure 1. All atoms are initially placed into a coarse grid. After the coarse grid FDPB calculation, the electrostatic energy of **1** is saved. Second, the electrostatic potentials of **2** and **3** are recalculated using a fine grid (focusing grid) with the initially calculated potential as the boundary condition. The total electrostatic energy of the complex is the summation of the calculated energies of **1**, **2**, and **3**. The identical procedure can be applied to proteins and ligands separately (i.e., free state). Since ligands are usually small, the calculations can be done using a fine grid directly.

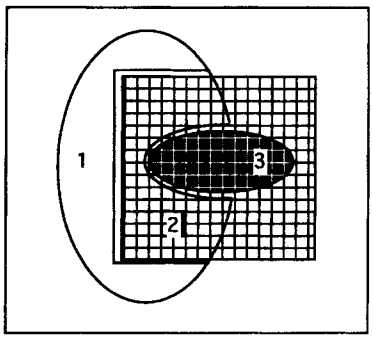

FIGURE 1. A schematic view of the FDPB calculation system. Regions **1** and **2** represent a protein and **3** a ligand. Initially, **1**, **2**, and **3** are placed into a coarse grid (outer square) to calculate the potential of the whole system. The potentials of **2** and **3** are then recalculated using a fine grid (inner meshed square) with the coarse grid potential as the boundary condition.

The FDPB calculation is carried out using the program UHBD[14] (Molecular Simulations Inc.) on Silicon Graphics computers. The computing setup specifications are summarized in Table I. The thermolysin with inhibitors Cbz-Gly-$\phi(PO_2)$-X-Leu-Leu or ZG^p(X)LL, where X = NH, O, or CH_2, is used as the test system.[15–19] All structures (PDB5TMN)[17] and force field parameters (GROMOS based) are identical to those in the previous calculation.[6]

ERROR ANALYSIS

The intrinsic error associated with the grids in FDPB calculation comes from the approximate charge (q) distribution and boundary dielectric (ε_b) distribution. In the program UHBD, these two inputs are assigned through the following two equations:

$$q = q_0(1 - a)(1 - b)(1 - c)$$

and

$$\varepsilon_b = \varepsilon_1\varepsilon_r/(\varepsilon_1(1 - a) + \varepsilon_r a)$$

where a, b and c are the fractional distances along the grid axes that the actual location of an atomic charge q_0 or molecular surface is from the grid point; and ε_1 and ε_r signify two different dielectrics separated by the molecular surface. The second equation actually is an improved dielectric assignment which smoothes the dielectric boundary. Obviously, any change in a, b and c affects q and ε_b, which in turn affects the potential through the PB equation. Although other assignment meth-

TABLE I.
Electrostatic energy calculation setup of UHBD.

Optional input	Coarse grid	Fine grid
Grid spacing (Å)	1.0	0.25
Grid dimension	100^3	100^3
Boundary condition	Debye-Huckel	Coarse grid potential
Ionic strength (mM)	100	100
Radius of the ions (Å)	1.4	1.4
Radius of surface probe (Å)	—	1.1
Number of surface points per atom	—	400
Smooth boundary dielectric	No	Yes

ods may improve the accuracy, the error produced seems inevitable as long as we use the grid representation for molecules.

Two types of errors associated with the grid in the binding energy calculation are analyzed. The statistical error is estimated by shifting molecular coordinates relative to the origin of the grid. To keep the cube-shaped system (2 + 3) the same in each calculation, the transitional averaging rather than the rotational averaging[11] is applied. Both averagings have the same effect of redistributing the charge and dielectric among the grid points. The systematic error due to the grid is examined by varying the spacing from 0.64 to 0.20 Å. Because reducing the grid spacing is limited by the array dimension and the size of a focusing system, a linear extrapolation is used to access the systematic error.

Results and Discussion

INTRINSIC ERRORS DUE TO GRIDS

Table II shows the statistical analysis of calculated solvation energies $\Delta\Delta G_{ele}$. Shifting solute coordinates relative to the grid produces very small ($\sim 0.1\%$) errors in solvation energies at the grid spacing of 0.25 Å. Due to the great solvation energies of proteins, the root mean square (rms) of solvation energies for bound states are on the order of 1 kcal/mol. If these errors were accumulated in $\Delta\Delta G_{ele}$, the calculated value could have unacceptable uncertainty. Fortunately, these statistical errors are mostly canceled in computing $\Delta\Delta G_{ele}$. In fact, the rms deviation of $\Delta\Delta G_{ele}$ is 0.03 kcal/mol, which is smaller than those of the solvation energies. Averaging over different grid spacing (less than 4.5 Å) yields similar rms deviation. Therefore, at small grid spacing, the statistical error is so small that it can be neglected when comparing calculated results with experimental values.

The $\Delta\Delta G_{ele}$ dependency of grid spacing can be seen clearly in Figure 2. Reducing grid spacing generally increases the absolute value of $\Delta\Delta G_{ele}$. The extrapolation yields 4.25 kcal/mol of the binding energy difference at zero spacing. Therefore, the estimated systematic error in $\Delta\Delta G_{ele}$ is about 0.1 kcal/mol at the grid spacing of 0.25 Å. This experiment also contains statistical errors indicated by the fluctuation of the curves in Figure 2. At small spacing, these statistical errors are on the same order as the one at the spacing of 0.25 Å. The errors become large when the grid spacing is greater than 0.5 Å.

It has been recommended that a grid spacing of $R/3$ be used in FDPB calculation,[7] where R is the van der Waals radius. The smallest van der Waals radius in the test system is 1.2 Å for hydrogen. Thus, a spacing of 0.4 Å or less is required. A second choice of the grid spacing can be related to the resolution of point charge. To avoid any two atomic charges partitioning into a common grid point, at least two grid points should exist along one of the Cartesian coordinates between the two atoms. For the smallest bond length (1 Å) in molecular systems, a grid spacing of 0.28 Å can satisfy this criterion. Further reducing grid spacing

TABLE II.
Statistics of calculated solvation energies and binding energy differences of thermolysin-inhibitor systems (kcal / mol).[a]

	ZGPLL		ZGP(O)LL		
	$\Delta G_{s,b}$	$\Delta G_{s,f}$	$\Delta G_{s,b}$	$\Delta G_{s,f}$	$\Delta\Delta G_{calc}$
Shift average[b]	−1365.788	−71.434	−1366.052	−69.687	4.110
rms Deviation	1.498	0.064	1.498	0.083	0.034
Relative error (%)	0.11	0.09	0.11	0.12	0.83
Grid average[c]	−1366.665	−71.648	−1366.924	−69.884	4.096
rms Deviation	3.406	0.340	3.404	0.302	0.041
Relative error (%)	0.25	0.47	0.25	0.43	1.00

[a] Done using Excel (Microsoft).
[b] A total of seven data sets at grid spacing of 0.25 Å were produced by shifting the origin of the grid by 0.1 and 0.2 Å along the x, y and z axes, respectively.
[c] A total of 15 data sets from grid spacing of 0.2 to 0.44 Å were averaged.

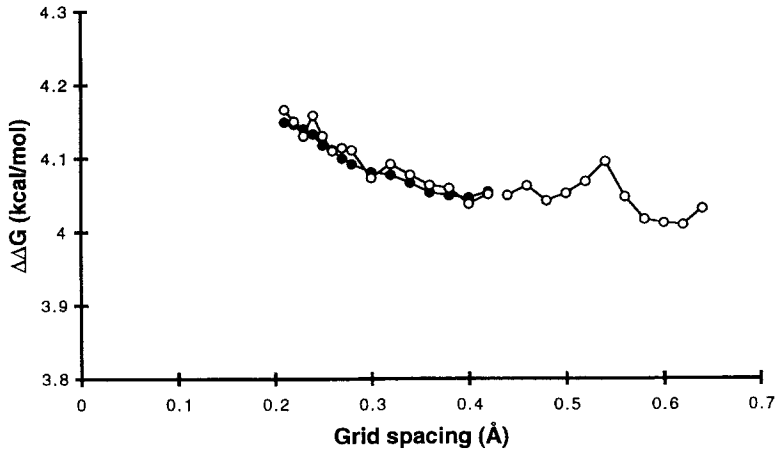

FIGURE 2. The calculated electrostatic binding energy difference (○) of the thermolysin-inhibitor system vs. grid spacing. The data at small spacing are averaged over three adjacent points to reduce the noise, shown as ●. A linear extrapolation of the averaged data yields 4.25 kcal / mol at the grid spacing of zero. All data processing and plotting, and thereafter, were done using Excel (Microsoft).

certainly improves the detailed description of charge distribution and solute–solvent interface. With the same array dimensions, however, the accuracy could be offset by using smaller spacing because of the exclusion of more atoms in the focusing grid.

EFFECTS OF SOLVENT PARAMETERS

Figures 3–5 show how the radius of a solvent probe, the number of surface points describing the solute–solvent interface, and ionic strength affect calculated solvation energies and $\Delta\Delta G_{ele}$. The functions of these parameters have been previously described by Gilson et al.[11] and were implemented in the UHBD program as user's options. Although all parameters affect solvation energies, only the radius of the solvent probe has noticeable effects on $\Delta\Delta G_s$'s which yield an energy difference of 0.3 kcal/mol between calculations using a radius of 1 Å and those using 1.4 Å.

Usually, a water probe of 1.4 Å is used to sample the solvent region in molecular modeling. However, a water pocket inside a protein could be

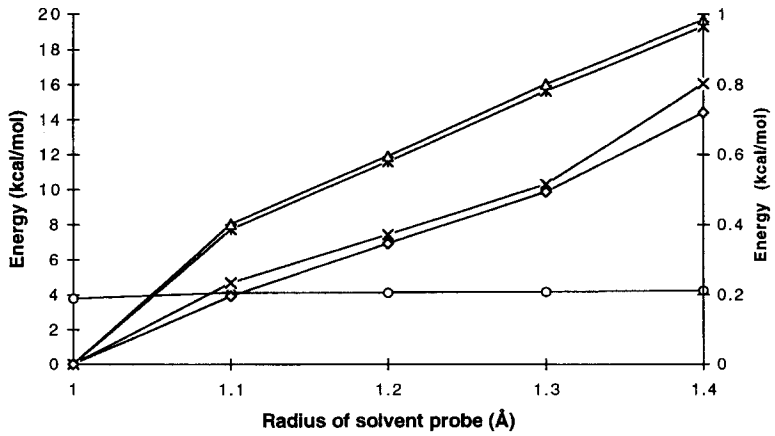

FIGURE 3. The effects of radius of solvent probe on ΔG_{ele} (●), and the relative values of $\Delta G_{s,b}$ (∗ and Δ) and $\Delta G_{s,f}$ (◇ and X, right axis) for ZGLL and ZG(O)LL, respectively. The variations of solvation energies are on the order of 1 kcal / mol for inhibitors and 10 kcal / mol for thermolysin. In contrast, the variation of binding energies is less than 0.2 kcal / mol.

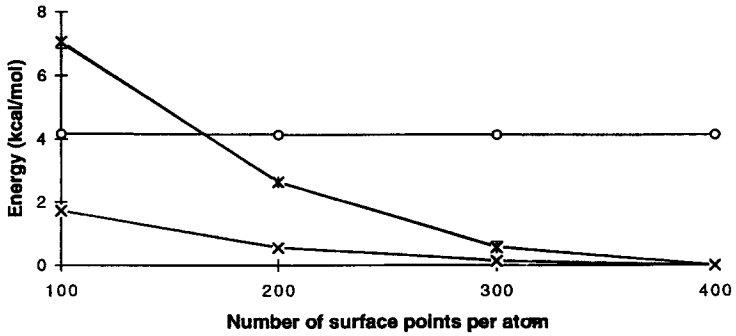

FIGURE 4. The effects of number of surface points on $\Delta\Delta G_{ele}$ (○), and the relative values of $\Delta G_{s,f}$ (X) and $\Delta G_{s,b}$ (*) for the ZG(O)LL system. The corresponding values of $\Delta G_{s,f}$ and $\Delta G_{s,b}$ for the ZGLL system are essentially identical to those of the ZG(O)LL system. The differences are too small to display in this plot. Increasing the number of surface points reduces the solvation energies while the binding energy differences remain constant.

assigned as a part of solute because none of enclosed grid points has at least a distance of 1.4 Å to the surface of a protein. To avoid this problem, a small probe of 1–1.1 Å was used in previous calculations.[5,6] Two questions may be raised at this point. First, a real water-size cavity of a protein could be assigned as a high-dielectric water. By graphically displaying the dielectric map together with the X-ray structure of the protein, one can easily identify the region if it exists and correct it using a dummy atom. Second, a binding water interacting specifically with solute atoms is treated as an implicit high-dielectric solvent. One could use an explicit water model in this case as long as the coordinates of water hydrogens are determined. The effect on the binding energy calculation using an explicit water has not been studied. There is an additional complication of how to treat the same water in the free state. On the other hand, the high-dielectric character of ice[21] experimentally supports using the implicit water model for crystallographically determined water in FDPB calculations.

The atomic surface points are used to determine more accurate dielectrics of the system. As shown in Figure 4, despite the large downshift of solvation energies with increased point density, the binding energies remain constant. The choice of

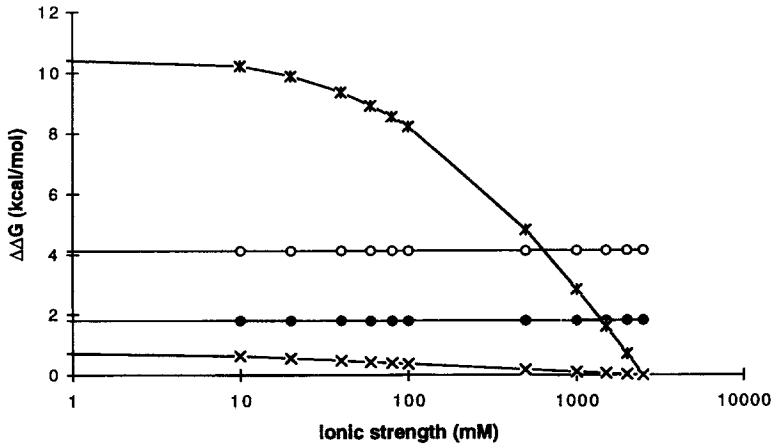

FIGURE 5. The effects of ionic strength on $\Delta\Delta G_{ele}$ [○ for ZG(O)LL and ● for ZG(C)LL, respectively] and the relative values of $\Delta G_{s,f}$ (X) and $\Delta G_{s,b}$ (*) for ZGLL. The relative values of $\Delta G_{s,f}$ and $\Delta G_{s,b}$ for ZG(O)LL and ZG(C)LL are not shown because they are not distinguishable from those of ZGLL on the scale. The variations of solvation energies are on the order of 1 kcal / mol and 10 kcal / mol for the inhibitors and thermolysin, respectively. $\Delta\Delta G_{ele}$ remains constant.

200 points per atom (default) seems suitable for the calculation without the loss of accuracy.

It is well known that increasing salt concentration reduces the electrostatic free energy of protein ions based on both PB equation and experimental observations.[22] The solvation energy (which differs from the electrostatic free energy by a gas phase electrostatic energy) of the testing molecules showed similar trends without exception (see Fig. 5). However, the variations in binding energy difference across three orders of magnitude of ionic strength are almost undetectable. Clearly, the variations in solvation energies are mostly canceled when computing $\Delta\Delta G_{ele}$. In practice, one need choose an ionic strength according to the experimental condition, and particularly when comparing binding energies derived from different labs. The insensitivity of the ionic strength as well as other solvent parameters tested in this work may be the consequence of these molecular modifications that involve few changes in solvent–solute interface. More pronounced effects could be observed if a part of solute is changed to solvent, or vice versa.

EFFECTS OF SOLUTE PARAMETERS

Figure 6 shows how the dielectric constant (D) of solute molecules affects $\Delta\Delta G_{ele}$. $\Delta\Delta G_{ele}$ is approximately scaled by the reciprocal of the dielectric constant. The shielding effect of solute dielectric on $\Delta\Delta G_{ele}$ results from the similar effects on ΔG_a and ΔG_s. The former is explicitly expressed as $1/D$ in Coulomb's law. The scaling factor in ΔG_s is not so obvious from the PB equation. Because the PB method can accurately reproduce the Born solvation energy, the scaling factor should be close to $(1/D - 1/\varepsilon)$ for a point charge, as expressed in the Born equation.[20] Indeed, all our calculations show that the effective scaling factor for the solvation energies of protein and ligand is slightly smaller that $1/D$.

As an empirical parameter, the solute dielectrics used in FDPB calculations are usually among 2 to 4 based on experimental dielectric constants for most organ molecules as well as the polarizability of proteins. Because the dielectric of 3 has been proven to be suitable for binding energy calculations with the current charge set, it is used throughout this work. Figure 6 also shows that changing the solute dielectric will not alter the binding affinity trend.

The results of calculations using two different sets of partial charges for inhibitors are listed in Table III. The partial atomic charge probably is one of the most sensitive parameters to the calculated energies. Our results show that a qualitatively correct charge set should be able to match experimental binding trends if the binding is dominated by electrostatics. The electrostatic potential fitted charge works better than those of Mulliken population in this work. A related issue is the assignment of protonation states for residues around a binding site. A unit charge could affect calculated results significantly. In fact, we incorrectly used default unprotonated Glu 143 and His 231 in our early work. Although the calculated binding trends are reasonable, the discrepancies between the calculation and experiment were as large as 1.5 kcal/mol.

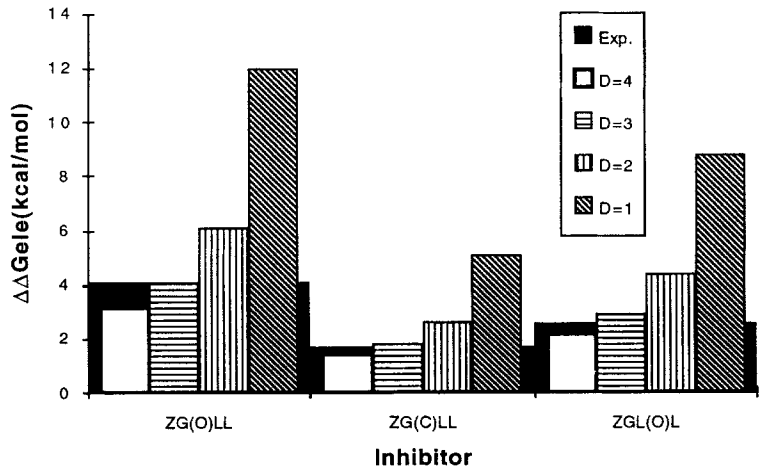

FIGURE 6. The effects of solute dielectric on $\Delta\Delta G_{ele}$. The heights of the black background represents experimental binding energies for the thermolysin system. The use of dielectric of 3 closely matches the experimental values.

TABLE III.
Calculated binding energy differences using Mulliken charges and electrostatic potential-fitted charges (kcal / mol).[a]

Modification	Mulliken	*E* Potential	Experiment[b]
NH to O	4.39	4.14	4.13
NH to CH_2	0.80	1.79	1.7, 0.1
NH′ to O′	3.10	2.76	2.58

[a] Partial charges were calculated using Spartan (wave function) with 6-31G* basis sets. For details, see ref. 6.
[b] From ref. 16.

Finally, the way of partitioning atomic charges into the grid could improve the precision of calculated binding energy further. The current trilinear method[14] can only give two effective digits after the decimal for protein solvation energies due to the large self-energy. The new uniform charge model[13] is expected to output at least three digits after the decimal. Although this precision may not seem necessary for the binding energy, the uniform method seems to be a better way to charge the grid in FDPB calculation. The uncertainty in $\Delta\Delta G_{ele}$ using this model should be similar to our results based on the reported uncertainties in solvation energies.[13] By the time we finished this work, the uniform charge model has not been implemented in UHBD.

Conclusion

We have examined the intrinsic error in binding energy calculation using the FDPB method. Following our procedure, the statistical and systematic errors due to discrete grid representation of molecular shape and charges amount to about 1% and 5% of the calculated binding energy differences, respectively. The greater accuracy results from a three-stage error cancellation: first in ΔG_s, then $\Delta\Delta G_s$, and finally $\Delta\Delta G_{ele}$. In addition, the recently developed techniques, such as focusing and smoothing, also contribute to the improvement. Among the parameters examined, the partial charge, dielectric constant, and radius of solvent can influence the calculated results. Whereas these problems may not be unique to the FDPB method, it is now feasible to optimize these parameters for application in a more realistic environment.

Acknowledgments

This work was completed when both authors were with Sterling Winthrop Inc., Collegeville, PA.

References

1. J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, UK, 1987.
2. C. L. Brooks III, M. Karplus, and B. M. Pettitt, *Adv. Chem. Phys.*, **71**, 59 (1988).
3. P. L. Cummins and J. E. Gready, *J. Computer-Aided Mol. Design*, **7**, 535 (1993).
4. Yun-yu Shi, A. E. Mark, Cun-xin Wang, F. Huang, H. J. C. Berendsen, and W. F. van Gunsteren, *Protein Engineer.*, **6**, 289 (1993).
5. J. Shen and F. A. Quiocho, *J. Comput. Chem.*, **16**, 445 (1995).
6. J. Shen and J. Wendoloski, *Protein Sci.*, **4**, 373 (1995).
7. M. K. Gilson and B. Honig, *Proteins*, **4**, 7 (1988).
8. S. C. Harvey, *Proteins*, **5**, 78 (1989).
9. M. E. Davis and J. A. McCammon, *Chem. Rev.*, **90**, 509 (1990).
10. K. Sharp and B. Honig, *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 301 (1990).
11. M. K. Gibson, K. Sharp, and B. Honig, *J. Comput. Chem.*, **9**, 327 (1987).
12. M. E. Davis and J. A. McCammon, *J. Comput. Chem.*, **12**, 909 (1991).
13. R. E. Bruccoleri, *J. Comput. Chem.*, **14**, 1417 (1993).
14. M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon, *Comp. Phys. Comm.*, **62**, 187 (1991).
15. P. A. Bartlett and C. K. Marlowe, *Science*, **235**, 569 (1987).
16. B. P. Morgan, J. M. Scholtz, M. D. Ballinger, I. D. Zipkin, and P. A. Bartlett, *J. Am. Chem. Soc.*, **113**, 297 (1991).
17. H. M. Holden, D. E. Tronrud, A. F. Monzingo, L. H. Weaver, and B. W. Matthews, *Biochemistry*, **26**, 8542 (1987).
18. D. E. Tronrud, H. M. Holden, and B. W. Matthews, *Science*, **235**, 571 (1987).
19. B. W. Matthews, *Acc. Chem. Res.*, **21**, 333 (1988).
20. M. Born, *Zeit. Phys.*, **1**, 45 (1920).
21. D. R. Lide, Ed., *CRC Handbook of Chemistry and Physics*, 75th ed., CRC Press, Boca Raton, FL, 1994.
22. C. Tanford, *J. Phys. Chem.*, **59**, 788 (1955).